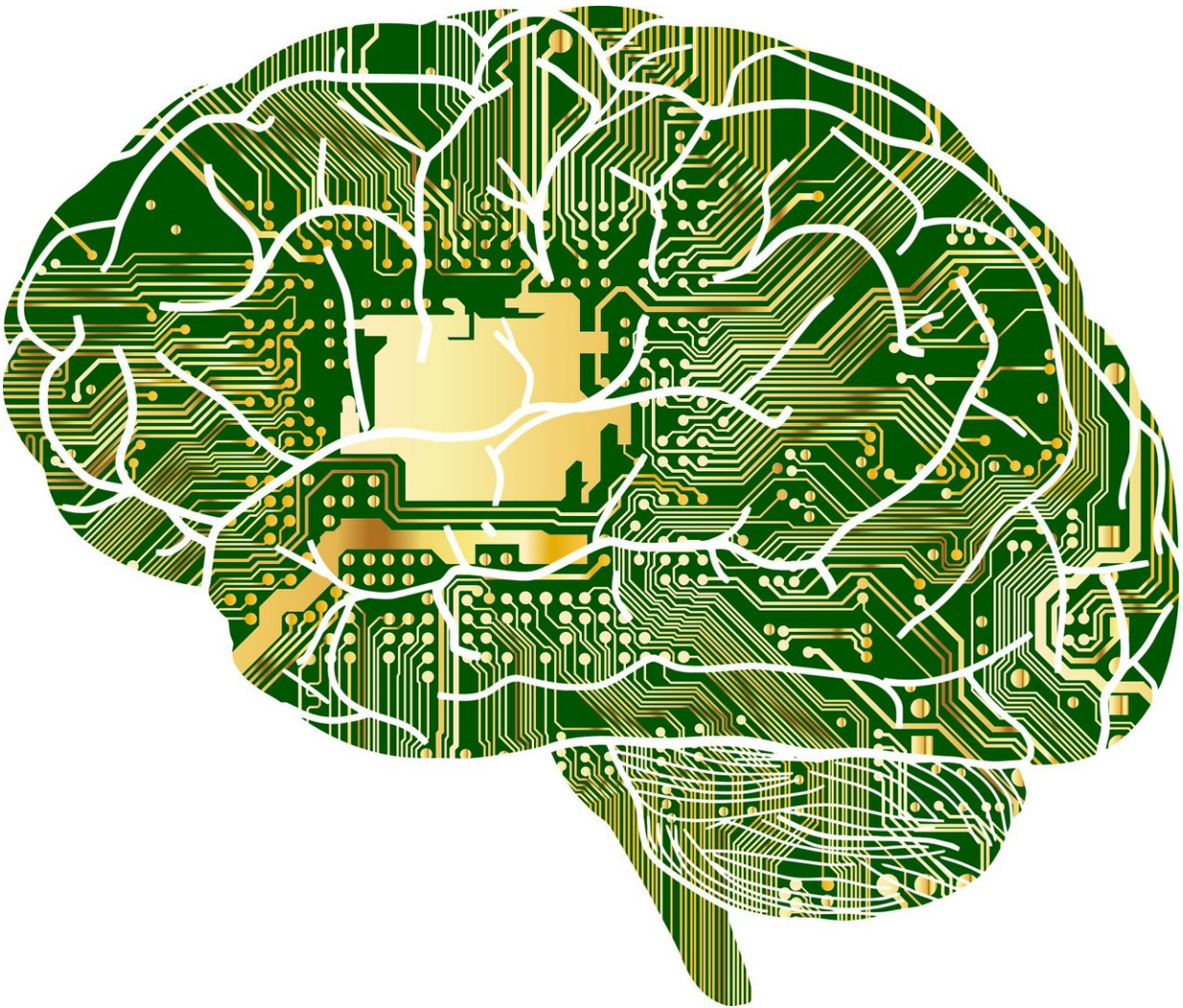# Artificial intelligence trained to analyze causation

June 6 2018

Credit: CC0 Public Domain

The causes of real-world problems in economics and public health can be notoriously hard to determine. Often, multiple causes are suspected, but large datasets with time-sequenced data are not available. Previous models could not reliably analyze these challenges. Now, researchers have tested the first artificial intelligence model to identify and rank many causes in real-world problems without time-sequenced data, using a multi-nodal causal structure and Directed Acyclic Graphs.

When something bad happens, it is natural to try figure out why it happened. What caused it? If the cause is determined, it may be possible to avoid the same outcome the next time. However, some of the ways in which humans try to understand events, such as resorting to superstition, cannot explain what is actually going on. Neither does correlation, which can only say that event B happened around the same time as event A.

To really know what caused an event, we need to look at causality—how information flows from one event to another. It is the information flow that shows there is a causal link—that event A caused event B. But what happens when the time-sequenced information flow from event A to event B is missing? General causality is required to identify the causes.

Mathematical models for general causality have been very limited, working for up to two causes. Now, in an artificial intelligence breakthrough, researchers have developed the first robust model for general causality that identifies multiple causal connections without time-sequence data, the Multivariate Additive Noise Model (MANM).

Researchers from the University of Johannesburg, South Africa, and National Institute of Technology Rourkela, India, developed the model and tested it on simulated, real-world datasets. The research is published in the journal *Neural Networks*.

"Uniquely, the model can identify multiple, hierarchical causal factors. It

works even if data with time sequencing is not available. The model creates significant opportunities to analyse complex phenomena in areas such as economics, disease outbreaks, climate change and conservation," says Prof Tshilidzi Marwala, a professor of artificial intelligence, and global AI and economics expert at the University of Johannesburg, South Africa.

"The model is especially useful at the regional, national or global level where no controlled or natural experiments are possible," adds Marwala.

Superstition and correlation towards causality

"If a black cat runs across the road, or an owl hoots on a roof, some people are convinced something really bad is going to happen. A person can think there is a connection between seeing the cat or the owl and what happened afterwards. However, from an artificial intelligence point of view, we say there are no causal links between the cat, the owl, and what happens to the people who see them. The cat or the owl were seen just before the event, but they are merely correlated in time with what happened later," says Prof Marwala.

Meanwhile, inside the house where the owl was sighted, something more sinister may be going on. The family inside may be sliding deeper and deeper into debt. Such a financial situation can impose severe restrictions on the household, eventually becoming a trap from which there is little escape. But do the people living there understand the actual causal connections between what happens to them, what they do, and their levels of debt?

Causality at household level

The causes of persistent household debt are a good example of what the new model is capable of, says post-doctoral researcher Dr. Pramod

Kumar Parida, lead author of the research article.

"At a household level one can ask: Has the household lost some or all of its income? Are some or all members spending beyond their income? Has something happened to household members that is forcing huge spend, such as medical or disability bills? Are they using up their savings or investments, which have now run out? Is a combination of these things happening, if so, which are the more dominant causes of the debt?"

If enough data about the household's financial transactions is available, complete with date and time information, it is possible for someone to figure out the actual causal connections between income, spend, savings, investments and debt.

In this case, simple causality theory is sufficient to find out why this household is struggling.

General causality at societal level

But, says Parida, "What are the real reasons most people in a city or a region are struggling financially? Why are they not getting out of debt?" Now, it is no longer possible for a team of people to figure this out from available data, and a whole new mathematical challenge opens up.

"Especially if you want the actual causal connections between household income, spend, savings and debt for the city or region, rather than expert guesses or 'what most people believe,'" he adds.

"Here, causality theory fails, because the financial transaction data for households in the city or region will be incomplete. Also, date and time information will be missing for some data. Financial struggle in low, middle and high-income households may be very different, so you'll

want to see the different causes from the analysis," says Parida.

"With this model, you can identify can identify multiple major driving factors causing the household debt. In the model, we call these factors the independent parent causal connections. You can also see which causal connections are more dominant than the others. With a second pass through the data, you can also see the minor driving factors, what we call the independent child causal connections. In this way, it is possible to identify a possible hierarchy of causal connections."

Significantly improved causal analysis

The Multivariate Additive Noise Model (MANM) provides significantly better causal analysis on real-world datasets than industry-standard models currently in use, says co-author Prof Snehashish Chakraverty, at the Applied Mathematics Group, Department of Mathematics, National Institute of Technology Rourkela, India.

"In order to improve a complex regional problem such as household debt or healthcare challenges, it may not be sufficient to have the knowledge of patterns of the debt, or of disease and the exposure. On the contrary, we should understand why such patterns exist, to have the best way of changing them. Previous models developed by researchers worked with a maximum of two causal factors, that is they were bivariate models, which simply could not find multiple feature dependency criteria," he says.

Directed Acyclic Graphs

"MANM is based on Directed Acyclic Graphs (DAGs), which can identify a multi-nodal causal structure. MANM can estimate every possible causal direction in complex feature sets, with no missing or wrong directions."

The use of DAGs is a key reason MANM significantly outperforms models previously developed by others, which were based on Independent Component Analysis (ICA), such as Linear Non-Gaussian Acyclic Model (ICA-LiNGAM), Greedy DAG Search (GDS) and Regression with Sub-sequent Independent Test (RESIT), he says.

"Another key feature of MANM is the proposed Causal Influence Factor (CIF), for the successful discovery of causal directions in the multivariate system. The CIF score provides a reliable indicator of the quality of the casual inference, which enables avoiding most of the missing or wrong directions in the resulting causal structure," concludes Chakraverty.

Where an existing dataset is available, MANM now makes it possible to identify multiple multi-nodal causal structures within the set. As an example, MANM can identify the multiple causes of persistent household debt for low, middle and high-income households in a region.

More information: Pramod Kumar Parida et al, A multivariate additive noise model for complete causal discovery, *Neural Networks* (2018). DOI: 10.1016/j.neunet.2018.03.013

Provided by University of Johannesburg